Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Forecasting Analysis of Heavy Metal Concentrations in the Ciliwung River Using Random Forest and Xgboost Methods

Lina Gozali

linag@ft.untar.ac.id

Tarumanagara University

Christyfianie Elva Angelica

Tarumanagara University

Adianto Adianto

Tarumanagara University

Togar Simatupang

Bandung Institute of Technology

Jose Arturo Garza-Reyes

University of Derby

Vikas Kumar

University of Portsmouth

Wan Hee Cheng

INTI International University

Research Article

Keywords: Ciliwung River, forecasting, heavy metal, random forest, XGBoost

Posted Date: October 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10855409/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

The Ciliwung River plays a crucial role as a water resource in the Jakarta Metropolitan area; however, industrial activities and urbanisation have increased the risk of heavy metal pollution, including Pb, Cd, and Hg. This study aims to: (1) compare the performance of Random Forest and Extreme Gradient Boosting (XGBoost) algorithms in predicting heavy metal levels, (2) forecast heavy metal concentrations for the period 2024–2030, and (3) evaluate the prediction results against the quality standards set out in Government Regulation of the Republic of Indonesia Number 22 of 2021. The model was trained using secondary water quality data (pH, BOD, COD, TDS, and TSS) and evaluated using the metrics MAE, MAPE, MSE, RMSE, and R^2. The results show that XGBoost performs better than Random Forest, with higher R^2 values for all parameters (Pb = 0.423; Cd = 0.553; Hg = 0.584). The projected heavy metal concentrations range from Pb (14.285–19.483 µg/L), Cd (7.413–9.998 µg/L), and Hg (0.227–0.457 µg/L). All prediction results are below the Class II threshold according to Government Regulation No. 22 of 2021 (Pb = 30 µg/L; Cd = 10 µg/L; Hg = 2 µg/L). Although Cd and Pb tend to approach the threshold values, regular monitoring is still necessary.

INTRODUCTION

Rivers serve as vital components of aquatic ecosystems, providing essential resources for human, animal, and plant life. However, rapid industrialisation and urban expansion have led to significant deterioration in river water quality, particularly due to heavy metal contamination [1], [2]. According to the World Health Organisation (WHO), heavy metal pollution poses a primary global concern due to its high toxicity [3]. Such contamination often originates from anthropogenic activities, including manufacturing industries, mining, modern agriculture, and the discharge of domestic waste [4], [5], [6]. Due to their persistent, toxic, and bioaccumulative nature, heavy metals persist in environmental media, including soil, water, and the atmosphere, for prolonged periods and may gradually reach hazardous concentrations [7], [8], [9], [10].

The accumulation of these pollutants severely threatens soil ecological functions, degrades water quality, and adversely impacts aquatic ecosystems. Furthermore, heavy metals can infiltrate food chains, causing chronic exposure in humans and increasing the risk of toxic and potentially carcinogenic effects. Metals such as mercury (Hg), lead (Pb), cadmium (Cd), chromium (Cr), copper (Cu), nickel (Ni), and zinc (Zn) are among the most concerning environmental pollutants due to their non-biodegradable characteristics and long-term accumulation in the environment [4], [11], [12], [13]. While some contamination arises from natural sources, including atmospheric deposition and geochemical interactions, anthropogenic sources, such as rapid urbanisation, industrial discharge, and mining, are the dominant contributors [14], [15], [16].

The Ciliwung River is a major, highly urbanised watershed that flows from the highlands of Bogor through Depok into the densely populated metropolitan core of Jakarta before discharging to Jakarta Bay. As one of the major rivers flowing through the Jakarta Metropolitan Area, the Ciliwung plays an essential

role in supplying water to local communities [17], [18], [19]. However, increasing population and industrial activity along its course have contributed to significant pollution, including the accumulation of heavy metals [20]. This condition not only threatens aquatic biodiversity but also poses risks to public health. Existing mitigation strategies remain limited by outdated monitoring practices. This condition aligns with Government Regulation of the Republic of Indonesia No. 22 of 2021 on the Implementation of Environmental Protection and Management, which mandates that water bodies comply with environmental quality standards, includingthe water quality standard, as benchmarks or evaluating and managing water pollution [21].

Conventional laboratory-based approaches for heavy metal analysis, though accurate, are time-consuming, costly, and labour-intensive, often requiring extensive field sampling campaigns [22], [13]. These reactive methods may delay mitigation efforts until after contamination has occurred. As environmental data becomes increasingly complex, conventional techniques are deemed inadequate, underscoring the need for advanced, data-driven methods. Therefore, the application of machine learning models such as Random Forests and Extreme Gradient Boosting (XGBoost) offers a promising alternative for predictive water-quality analysis.

Recent studies have demonstrated the increasing application of machine learning techniques in environmental quality prediction. Bhagat et al. (2021) developed an XGBoost-based SuperLearning approach to predict Pb contamination in bay sediments using historical data from Bramble Bay and Deception Bay, Australia. Their study compared XGBoost with Random Forests, Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs), with XGBoost achieving the highest predictive performance [22]. Sharma et al. (2024) applied Random Forest, Decision Tree, and K-Nearest Neighbours (KNN) algorithms to predict water quality using physicochemical parameters, including pH, turbidity, dissolved oxygen, temperature, phosphates, and nitrates. Random Forest achieved the highest classification accuracy among the evaluated models [23]. More recently, Proshad et al. (2025) integrated machine learning algorithms, including XGBoost, LightGBM, CatBoost, and Gradient Boosting, with outlier detection techniques to predict heavy metal contamination in soil. Their results demonstrated that combining XGBoost with DBSCAN improved the prediction of Cd contamination. These studies demonstrate the potential of machine learning to capture complex and non-linear relationships in environmental datasets [24].

Despite these developments, several gaps remain. Existing studies have predominantly focused on either water-quality classification or heavy-metal prediction in soil and sediment. At the same time, comparatively limited attention has been given to future forecasting of multiple heavy metals in river water. Furthermore, although individual studies have compared several machine learning algorithms, there remains a need for comparative evaluation of ensemble regression methods for predicting Pb, Cd, and Hg simultaneously under the specific conditions of an urban river system. The Ciliwung River presents an important case because it flows through densely populated, highly developed areas where anthropogenic activities may affect water quality. In addition, the availability of long-term monitoring

data provides an opportunity to investigate the application of machine learning for future heavy metal forecasting.

This study aims to utilise these models to forecast heavy metal concentrations in the Ciliwung River using historical water quality data from governmental agencies, such as the Environmental Agency and Statistics Indonesia (BPS), as well as relevant secondary sources. Specifically, the objectives of this study are: (1) to compare the performance of Random Forest and XGBoost in predicting heavy metal levels, (2) to forecast concentrations from 2024 to 2030, and (3) to evaluate the predicted values against the regulatory thresholds defined by Government Regulation of the Republic of Indonesia No. 22 of 2021. Through this approach, the study aims to contribute to more efficient, proactive, and data-driven river management strategies that support sustainable environmental governance.

LITERATURE REVIEWS

Within the broader framework of Environmental Science, both Water Quality Science and Aquatic Environmental Science represent key interdisciplinary branches that focus on understanding, monitoring, and managing water environments, particularly rivers, lakes, wetlands, and other freshwater systems [25].

Water Quality Science is a specialised field that examines the physical, chemical, and biological characteristics of water to evaluate its suitability for ecological functions, human consumption, agriculture, industry, and recreation. This discipline investigates pollutants such as nutrients, organic matter, pathogens, and heavy metals (e.g., Pb, Cd, Hg), as well as processes controlling contaminant transport, transformation, and removal in aquatic systems. Researchers in Water Quality Science develop monitoring strategies, predictive models, treatment technologies, and regulatory standards to protect public health and sustain the environment. The field integrates environmental chemistry, microbiology, hydrology, and data science, increasingly incorporating machine learning and real-time sensing technologies for proactive water management [26].

Aquatic Environmental Science, by contrast, adopts a broader ecosystem perspective. It studies the structure, function, and health of aquatic ecosystems, emphasising interactions among water quality, hydrological processes, sediments, aquatic organisms, and surrounding landscapes. Beyond pollutant measurement, this field evaluates ecological responses to environmental stressors, including pollution, climate change, land-use alteration, and urbanisation. Topics include ecosystem resilience, biodiversity conservation, bioaccumulation of contaminants, trophic interactions, and restoration of degraded river systems. Aquatic Environmental Science, therefore, links environmental processes with ecological outcomes and sustainable resource management [27].

Within the larger discipline of Environmental Science, these two areas complement one another. Environmental Science seeks to understand the relationships between natural systems and human activities to support sustainable development. Water Quality Science provides the analytical and technological foundation for detecting and controlling pollution. At the same time, Aquatic

Environmental Science supplies the ecosystem-based understanding necessary for long-term environmental protection and river basin governance. Together, they form an integrated scientific approach that moves environmental management from reactive pollution control toward preventive, data-driven, and ecosystem-centred water sustainability [28].

Bhagat et al. (2021) proposed an XGBoost-based SuperLearning model to predict Pb contamination in bay sediments using historical data from Bramble Bay and Deception Bay, Australia. The model was compared with Random Forests, Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs). XGBoost achieved the best performance, with an R^2 of 0.97, demonstrating its high accuracy and reliability for predicting heavy metal contamination [22].

Sharma et al. (2024) developed machine learning models to predict water quality using a dataset containing physicochemical parameters, including pH, turbidity, dissolved oxygen, temperature, phosphates, and nitrates. The performance of Random Forest, Decision Tree, and K-Nearest Neighbour (KNN) was compared. The results showed that Random Forest achieved the highest classification accuracy of 70.4%, outperforming the other models and demonstrating its effectiveness in handling complex, non-linear water quality data [23].

Proshad et al. (2025) developed machine learning models integrated with outlier detection techniques to predict heavy metal contamination in soil using data from Narayanganj, Bangladesh. Several algorithms, including XGBoost, LightGBM, CatBoost, and Gradient Boosting, were evaluated alongside outlier detection methods such as DBSCAN and Isolation Forest. The results showed that combining XGBoost with DBSCAN improved Cd prediction accuracy, increasing the R^2 value by 14.47% [24].

Heavy Metal

Heavy metals are metallic elements with a density greater than 5 g/cm³ that are toxic at specific concentrations and have a negative impact on the environment [29], [30]. Common environmental pollutants include Hg, Pb, Cd, Cr, Cu, Ni, and Zn. Sources of pollution can originate from natural processes, such as volcanic activity and rock weathering, as well as from anthropogenic sources, including industrial, mining, and domestic waste, and the use of fertilisers and pesticides [31], [32]. In aquatic ecosystems, heavy metals are dangerous because they are persistent and tend to bioaccumulate, which can ultimately affect human health through the consumption of contaminated biota [33], [34]. Acute exposure to Hg and Pb can cause abdominal colic, bloody diarrhoea, and kidney failure, while chronic exposure to Hg (especially methylmercury) causes nerve damage [35], [36], [37]. Pb is associated with neurological, respiratory, urinary, and cardiovascular disorders through oxidative stress and inflammatory mechanisms. Meanwhile, Cd also poses a high risk because it disrupts the body's physiological mechanisms, both in the short- and long-term [38].

In Indonesia, water quality control is regulated in Government Regulation No. 22 of 2021 concerning the Implementation of Environmental Protection and Management. This regulation classifies water bodies into four use classes: Class I for raw drinking water; Class II for water recreation, freshwater fish farming,

livestock farming, and irrigation; Class III for freshwater fish farming, livestock farming, and irrigation; and Class IV for agricultural irrigation. The threshold limits for heavy metal concentrations differ by Class, as shown in Table 1 [21].

Table 1
Water Quality Standards for Heavy Metals

Heavy Metal	Unit	Class I	Class II	Class III	Class IV
Lead (Pb)	μg/L	30	30	30	500
Cadmium (Cd)	μg/L	10	10	10	10
Mercury (Hg)	μg/L	1	2	2	5

These quality standards are an essential reference in this study, because the predicted heavy metal concentrations will be compared with applicable standards to assess the safety level of river water quality. Table 1 shows the concentration limits of heavy metals for water quality classification into four regulatory standard classes (Class I–IV), which generally represent various standards for permissible water use as drinking water supply, or water for recreation, water for fisheries and aquaculture industries, or water for agriculture, or water for industrial purposes. A more in-depth interpretation shows how environmental standards can achieve an integrated balance among human health protection, ecological security, and practical water use.

The data in the table show that Lead (Pb) has a permissible concentration of 30 µg/L in Class I–III waters, but increases sharply to 500 µg/L in Class IV waters. This pattern suggests that Classes I–III are intended for sensitive heavy metal uses—most likely as drinking water sources, aquatic ecosystems, or recreational waters—where sufficiently stringent protection against lead's neurotoxic effects is a primary concern. The significant relaxation of the limit in Class IV may reflect water bodies primarily intended for industrial or low-level contact use, where the risk of direct human exposure to this metal is minimal. These different regulatory standards highlight the strong toxicity and potential for bioaccumulation of lead, particularly the irreversible impacts on neurological development and certain aquatic organisms.

In contrast, cadmium (Cd) maintains a uniform limit of 10 µg/L across all classes. This consistency demonstrates that cadmium levels are considered highly hazardous even at very low concentrations, regardless of the classification of the water resource. Cadmium, its carcinogenicity, and its tendency to accumulate in humans require stringent control standards across all environmental contexts. The lack of relief in Class IV suggests that cadmium contamination poses a significant long-term ecological risk beyond direct human exposure, as it accumulates in sediments and biomagnifies in the food supply chain.

For mercury (Hg), standards have shown a gradual increase from 1 µg/L in Class I to 5 µg/L in Class IV, reflecting a tiered risk-management approach. Mercury is among the most toxic heavy metals because it can transform into methylmercury, which biomagnifies through the aquatic food chain. The stricter thresholds in Class I emphasise the protection of drinking water resources and sensitive ecosystems. In contrast, the slightly higher limits in the lower classes recognise exposure pathways for which the standards are of reduced value, while maintaining very tight controls compared to other metals.

Overall, Table 1 demonstrates a risk-based water quality classification system in which permissible pollutant concentrations are strictly regulated by standards that address water use and exposure risks. Heavy metals with severe toxicity and bioaccumulative characteristics (Cd and Hg) maintain consistently strict standard limits, while metals with more context-dependent exposure pathways (Pb) exhibit varying thresholds. This regulatory framework reflects the principles of modern environmental management: rather than imposing the same standards for all waters, policymakers implement preventative, proactive, and adaptive environmental governance that integrates ecological vulnerability, toxicology, and socio-economic water use.

Scientifically, the standards also guide control programs, water pollution control strategies, and the selection of water treatment technologies. Authorities responsible for controlling water classified under higher levels of protection require knowledge of advanced hazardous-metal removal technologies and ongoing monitoring. In comparison, those responsible for lower levels may prioritise reducing water pollution and gradual recovery. Therefore, Table 1 not only represents important regulatory standards but also illustrates how environmental science can translate toxicological knowledge into practical water resource management and sustainable environmental policies.

Forecasting

Forecasting is the process of predicting future conditions based on historical data, utilising mathematical models [38], [39]. The main principle is that past patterns and behaviour can be used to estimate future trends or anomalies. Generally, forecasting methods are categorised into two types: qualitative and quantitative [40]. Qualitative methods rely on subjective assessments or expert opinions, and are typically used when historical data is limited. In contrast, quantitative methods utilise statistical and mathematical approaches, such as time-series analysis and regression. The selection of the appropriate method is based on the lowest level of prediction error [41], [42].

Random Forest

Random Forest is a decision tree-based ensemble learning algorithm designed to enhance prediction accuracy while minimising the risk of overfitting [43]. This algorithm builds several decision trees using bootstrap sampling techniques and random feature subset selection at each node split [44]. Prediction results are obtained through a voting mechanism for classification or averaging for regression. The

advantages of Random Forests include their ability to handle high-dimensional data, resistance to outliers, and the ability to measure variable contributions through feature importance [45].

Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a decision tree-based machine learning algorithm that applies gradient boosting techniques to improve prediction accuracy [43]. This algorithm builds trees sequentially by minimising the loss function using gradient descent, and adds regularisation (L1 and L2) to reduce overfitting [28]. Shrinkage (learning rate) and subsampling mechanisms are also applied to improve model stability and generalisation. XGBoost excels in computational efficiency, can handle large and complex datasets, and is flexible with numerical, categorical, and missing values, making it widely used across applications such as water-quality prediction, classification, and risk analysis.

METHODS

The research methodology employed in this study utilised secondary data obtained from various credible sources. The research process followed a structured approach to ensure the accuracy and reliability of the findings [49]. The research methodology began with a literature review to gain a comprehensive understanding of the fundamental concepts of heavy metals, water pollution, and machine-learning-based forecasting approaches. The literature used included scientific publications, relevant agency reports, and government regulatory documents, including Government Regulation of the Republic of Indonesia Number 22 of 2021 concerning the Implementation of Environmental Protection and Management. The next stage is problem identification, focusing on heavy metal pollution in the Ciliwung River. The problem is formulated into a research topic with an emphasis on predicting the concentrations of the primary heavy metals, namely Pb, Cd, and Hg, so that this study is expected to contribute to strengthening the preventive water quality monitoring system. Furthermore, the research objectives were explicitly formulated, with a scope limited to the three types of heavy metals and a restriction to the Random Forest and XGBoost algorithms for the forecasting method.

The research data were obtained from the DKI Jakarta Provincial Environment Agency via a report on the 2024 water quality monitoring results for the Ciliwung River. The dataset used in this study comprises 324 data points from the River Water Quality Monitoring Reports for 2015 to 2023, covering 18 sampling locations. The data used consist of two main categories. First, data on heavy metal levels, including Pb, Cd, and Hg concentrations, were collected at various monitoring points in the Ciliwung River area, DKI Jakarta. Second, water quality parameters that affect pollution levels, namely pH, Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Total Dissolved Solids (TSS), and Total Suspended Solids (TDS), were used as the predictor variables. Climate or seasonal factors were not included because consistent historical data were unavailable for the entire study period. These data were used as input variables for the forecasting modelling process.

For historical monitoring data, missing values were estimated by interpolation from adjacent observations whenever sufficient temporal information was available. After preprocessing, the dataset was reviewed to confirm that no missing values remained before model training. Before forecasting the concentrations of Pb, Cd, and Hg for the 2024–2030 period, the future values of the predictor variables (pH, BOD, COD, TDS, and TSS) were estimated using a simulation approach based on the historical Compound Annual Growth Rate (CAGR) calculated from the available monitoring data collected between 2015 and 2023. The calculated CAGR for each predictor variable (pH, BOD, COD, TDS, and TSS) was subsequently applied to estimate annual values for the 2024–2030 period. To account for the inherent uncertainty and natural variability of future environmental conditions, a random variation of ± 2% was incorporated into the simulated predictor values. The ± 2% range was selected as a conservative sensitivity assumption to represent moderate year-to-year fluctuations without introducing unrealistic variability into the projected water quality parameters.

Modelling was performed using two machine learning algorithms: Random Forest and XGBoost. Random Forest is a decision tree-based ensemble method that builds several classification and regression trees and combines their predictions to achieve more stable accuracy. Meanwhile, XGBoost was applied as a boosting method that gradually combines weak models to reduce residual error, thereby producing a model with higher predictive accuracy. The dataset was divided into training and test data to ensure the validity of the prediction results. At the same time, the model performance was evaluated using various statistical metrics, namely Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2).

All data preprocessing, model development, cross-validation, and performance evaluation were conducted using Python 3.12.13 in the Google Colab environment. The main packages used were NumPy 2.0.2, Pandas 2.2.2, scikit-learn 1.6.1, and XGBoost 3.3.0. A fixed random seed of 42 was used to ensure reproducibility of the model development and evaluation procedures.

The next step is to analyse and interpret the results of heavy metal concentration predictions. The analysis examines trends in the concentration changes of Pb, Cd, and Hg from 2024 to 2030. Next, the prediction results are compared with the Class II river water quality standards as stipulated in Government Regulation No. 22 of 2021. This comparison aims to determine whether the heavy metal concentrations in the forecast period remain within permissible limits. Based on these results, this study draws conclusions that confirm the main findings and provides strategic recommendations to inform sustainable efforts to manage river water quality. The research methodology flowchart is presented in Fig. 1.

RESULT AND DATA ANALYSIS

In this study, the relationships between water quality variables (pH, BOD, COD, TDS, and TSS) and heavy metal concentrations (Pb, Cd, and Hg) were modelled using two algorithms: Random Forest and

XGBoost. The research data were randomly split into 80% for training and 20% for testing, using the random_state = 42 parameter to ensure reproducibility. Five-fold cross-validation was subsequently performed on the training set using shuffled folds with random_state = 42. The Random Forest model was built using the fit() function of RandomForestRegressor with default parameters, and then evaluated using MAE, MAPE, MSE, RMSE, and the coefficient of determination (R^2) metrics via the evaluate() function to assess the closeness of the predicted values to the actual values. Next, the trained model is used to project heavy metal levels for the 2024–2030 period, using simulated inputs based on the CAGR and a random variation of ± 2% to account for natural uncertainty. On the other hand, the XGBoost method is applied separately for each heavy metal target because the XGBRegressor library does not yet support multi-output regression. Therefore, three separate models are built for Pb, Cd, and Hg.

The Random Forest model was implemented with 100 trees (n_estimators = 100), random_state = 42, and parallel processing enabled (n_jobs = -1). Other hyperparameters, including max_depth, min_samples_split, and min_samples_leaf, were maintained at their default values. The XGBoost model was implemented with 100 boosting rounds (n_estimators = 100), the squared-error regression objective (objective = reg:squarederror), random_state = 42, and n_jobs = -1. Other hyperparameters, including learning_rate, max_depth, subsample, and colsample_bytree, were maintained at their default values.

In contrast, other parameters were kept at their default values to strike a balance between accuracy and prevent overfitting. The training process used the fit() function for each target, and the results were evaluated using the same metrics as for Random Forest. The trained XGBoost model was also used to project heavy metal levels for the 2024–2030 period, using inputs identical to those of the Random Forest simulation through the predict() function.

Based on the data obtained, calculations were performed to evaluate the Random Forest and XGBoost forecasting models in predicting the content of heavy metals lead (Pb), cadmium (Cd), and mercury (Hg). This calculation was done to evaluate the accuracy and reliability of the prediction models [46]. The evaluation was conducted using five metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results of the comparison of heavy metal content prediction models are presented in Table 2.

Table 2
Comparison of the Evaluation of the Heavy Metal Prediction Models

Methods	Heavy Metal	MAE	MAPE (%)	MSE	RMSE	R^2
Random Forest	Pb	3,302	16,656	11,947	3,456	0,274
	Cd	1,032	16,068	1,391	1,180	0,440
	Hg	0,050	26,802	0,003	0,058	0,375
XGBoost	Pb	2,908	14,539	9,49	3,081	0,423
	Cd	0,808	13,248	1,112	1,055	0,553
	Hg	0,045	22,469	0,002	0,047	0,584

Table 2 compares the predictive performance of the Random Forest and XGBoost methods for heavy metal concentrations (Pb, Cd, and Hg) using several statistical evaluation metrics, including MAE, MAPE, MSE, RMSE, and the coefficient of determination (R^2). In-depth analysis shows that these results not only reflect the model's accuracy, thereby reducing forecasting errors, but also demonstrate the method's ability to capture the complexity of heavy metal pollution dynamics in the Ciliwung River water system. Projected concentrations of heavy metals Pb, Cd, and Hg for the period 2024–2030, as determined using the XGBoost method, are presented in line chart form in Figs. 2 to 4.

In general, the XGBoost model performed better than Random Forest across all heavy metal parameter values. In Lead (Pb) metal forecasting, XGBoost produced a lower error value (MAE = 2.908; RMSE = 3.081) compared to Random Forest (MAE = 3.302; RMSE = 3.456), as well as an increase in the R^2 value from 0.274 to 0.423. These calculations show that XGBoost is better at explaining variations in Pb concentrations in the data. For overall environmental influences, Pb concentrations are usually influenced by highly dynamic processes such as existing industrial activities, parts of rainwater, melted ice, or irrigation that flow over the land surface into rivers, as well as sediment-water interactions, so the resulting model is based on gradient boosting, which can capture non-linear relationships and complex variable interactions more effectively.

For Cadmium (Cd), the model performance improvement was more significant. XGBoost produced an MAE of 0.808, an RMSE of 1.055, and an R^2 of 0.553, which is higher than that of Random Forest (R^2 = 0.440). An R^2 value above 0.5 indicates that more than half of the variation in Cd concentration can be explained by the prediction model. This result indicates that the Cd distribution pattern is relatively stable and strongly correlated with water quality parameters, such as pH, conductivity, and dissolved oxygen, making it easier for machine learning algorithms to learn.

The most obvious performance improvement occurred for Mercury (Hg). Despite the small absolute error (MAE) due to the naturally low Hg concentration, XGBoost significantly improved the model's explanatory power (R^2 = 0.584) compared to Random Forest (R^2 = 0.375). Mercury exhibits complex environmental behaviour due to chemical transformations, especially the formation of highly toxic

methylmercury. XGBoost's incremental error correction via boosting enables the model to capture small yet significant changes in environmental data, such as those observed in the Ciliwung River.

Overall, the decreases in MAE, MSE, and RMSE values obtained by XGBoost compared with Random Forest indicate that XGBoost produced relatively smaller prediction errors. In contrast, its lower MAPE values indicate better relative predictive performance. However, the MAPE value remaining above 10%—especially for Hg—indicates that prediction uncertainty remains, so machine learning models should be used as a monitoring support tool to complement data obtained from laboratory analysis, not as a complete replacement for chemical laboratory testing.

These results suggest that XGBoost can capture some non-linear relationships between the selected water quality parameters and heavy metal concentrations and may provide useful estimates to support environmental monitoring. However, the moderate R^2 values also indicate that the model does not fully represent the complex factors influencing Pb, Cd, and Hg concentrations in the Ciliwung River. Factors such as rainfall, river discharge, seasonal variation, industrial activities, and land-use characteristics were not incorporated into the model and may contribute to the unexplained variability. Therefore, the application of XGBoost should be considered a complementary approach to conventional water quality monitoring. Implementation of this model has the potential to improve the efficiency of water quality monitoring by enabling early detection of water pollution, optimising water data sampling frequency, and strengthening data-driven decision-making and policy-making systems for sustainable aquatic environmental management.

Figure 2 illustrates the predicted trend in Lead (Pb) concentrations from 2024 to 2030, compared with the established water quality standard of 30 µg/L (indicated by the red dotted line). Several important scientific and environmental interpretations can be drawn from this visualisation.

First, predicted Pb levels remain consistently well below regulatory thresholds throughout the projection period. Values range from 14 to 20 µg/L, indicating that, according to model predictions, water quality is consistently within acceptable safety limits. This result suggests that existing Pb pollution control and monitoring measures, natural dilution processes, or watershed management strategies may be effective in preventing critical levels of Pb contamination.

However, despite compliance with standards, the figure shows a gradual upward trend after 2026, peaking around 2029. This upward trend is environmentally significant because heavy metals, such as lead, exhibit cumulative and persistent behaviour in aquatic systems. Even concentrations below regulatory limits can pose long-term ecological and human health risks due to bioaccumulation in sediments and aquatic organisms. Therefore, the absence of regulatory violations does not necessarily imply zero environmental risk.

The observed fluctuations between years indicate that various environmental factors, including hydrological changes, urban runoff, variability in industrial discharges, and seasonal sediment mobilisation, influence Pb dynamics. The predicted increase by 2029 may signal increasing

anthropogenic pressure or a decline in the self-purification capacity of water bodies. From a management perspective, this pattern serves as an early warning indicator, emphasising the importance of preventive monitoring rather than reactive interventions.

Scientifically, the figures demonstrate the practical value of machine learning-based forecasting in environmental management. By projecting future contaminant behaviour relative to regulatory standards, decision-makers can anticipate potential risks before violations occur. The results suggest that, as long as water quality remains acceptable, proactive mitigation strategies, strengthened monitoring programs, and source control policies should be maintained to prevent long-term increases that approach or exceed quality limits.

In short, these figures indicate regulatory compliance but also the emergence of environmental pressures, highlighting that sustainable water governance should not only focus on meeting standards but also on maintaining long-term ecological resilience and preventing the gradual accumulation of heavy metals.

Figure 3 shows the predicted trend in cadmium (Cd) levels for the period 2024–2030 compared to the water quality standard of 10 µg/L. From a scientific and environmental management perspective, several important interpretations can be made.

First, Cd concentrations are generally very close to the standard threshold, ranging from 7.4 to 10 µg/L. Unlike Pb, which falls well below water quality standards, Cd levels exhibit a very narrow safety margin (tolerance value), indicating the Ciliwung River system is quite vulnerable. Predictions of near-consumption of the standard in 2027 and again in 2030 indicate that water quality is periodically at a critical level. This situation serves as a warning signal that even a small increase in pollutant load could potentially lead to exceeding permissible environmental standards.

Second, the observed fluctuation pattern reflects the high sensitivity of cadmium concentrations to environmental factors and anthropogenic activities. Cd levels in water are generally strongly influenced by variations in industrial discharges, agricultural runoff, mining residues, and changes in water chemistry, such as pH and redox conditions. The sharp decline in 2029 likely reflects natural dilution, improved pollution control, or specific hydrological conditions that reduce the availability of cadmium in the water column. However, the rebound in 2030 indicates that the primary source of pollution remains and has not been fully controlled. This condition requires high vigilance regarding Cd levels in the Ciliwung River.

From an ecological and public health perspective, cadmium is a highly toxic, bioaccumulative, and non-biodegradable heavy metal. Although Cd metal concentrations meet quality standards, long-term exposure at levels close to the threshold can potentially cause adverse accumulation in sediments and aquatic organisms and increase the risk of biomagnification in the food supply chain. Therefore, compliance with standards does not necessarily reflect a completely safe existing environmental condition.

Furthermore, this graph demonstrates the crucial role of early detection and prediction of pollution risks. The closeness of the predicted values to the quality standards underscores the need for a preventative approach to water management, such as strengthening pollution source control, increasing the frequency of water quality monitoring, and implementing more targeted and intensive water pollution mitigation strategies. Without adequate intervention, future environmental variability may lead to Cd concentrations exceeding permissible limits.

Overall, Fig. 3 shows that the cadmium concentration still meets water quality standards but is at a critical, unstable threshold and is not yet completely safe, so Cd can be categorised as a priority parameter that requires continuous monitoring and proactive environmental management to prevent future pollution.

Figure 4 shows the predicted trend in mercury (Hg) concentrations for the period 2024–2030, compared with the standard water quality limit of 2 μg/L. Based on this visualisation, several scientific analyses and environmental implications can be explained in depth.

First, all predicted Hg concentrations were well below the water quality standard threshold throughout the observation period. Hg concentrations are estimated to remain in the range of 0.22–0.46 μg/L, representing only 10–25% of the required water quality standards. This condition indicates that regulatory water quality remains within a relatively safe range and has not yet shown any critical risk of mercury contamination.

Second, the trend pattern shows a gradual decrease after 2026, reaching its lowest value in 2029, then a slight increase in 2030. This decrease can be interpreted as an indication of improved effectiveness of pollution source control, of natural dilution through heavy metal sedimentation, or of improvements in the aquatic environment. The small fluctuations observed reflect the natural dynamics of the aquatic system, influenced by hydrological variations, anthropogenic activities, and changes in water chemistry.

Nevertheless, ecologically, mercury remains a highly toxic heavy metal, primarily because it can convert to methylmercury, which readily accumulates in aquatic organisms and biomagnifies throughout the food supply chain. Therefore, even though concentrations remain well below quality standards, ongoing monitoring is necessary because the impacts of mercury are often long-term and cannot be inferred solely from water concentrations.Dilihat dari pelajaran manajemen lingkungan,

Figure 4 indicates that Hg pollution is relatively under control. However, the management approach should remain preventative, proactive, and adaptive through monitoring emission sources, controlling industrial and mining activities, and monitoring sediment quality and aquatic biota. This condition is crucial to maintaining the predicted stable, low Hg levels in the Ciliwung River in the long term.

Overall, Fig. 4 shows that mercury concentrations are at safe, stable, and controlled levels, and there is a decreasing trend in the risk of contamination. These predictions indicate a good level of environmental

resilience to Hg contamination and emphasise the importance of a continuous monitoring system to maintain the quality of aquatic ecosystems in the future.

The predictive performance observed in this study is generally consistent with previous research demonstrating the effectiveness of ensemble learning algorithms for environmental prediction. Bhagat et al. (2021) reported that XGBoost outperformed Random Forest, Artificial Neural Networks (ANN), and Support Vector Machines (SVM) in predicting Pb contamination in bay sediments, achieving an R^2 value of 0.97 [22]. Similarly, Proshad et al. (2025) found that XGBoost, when combined with DBSCAN outlier detection, improved the prediction accuracy of soil Cd contamination by increasing the R^2 value by 14.47% [24]. These findings support XGBoost's superior capability for modelling complex non-linear relationships in environmental datasets, consistent with the results of the present study, which found that XGBoost also outperformed Random Forests based on lower prediction errors and higher R^2 values.

However, the R^2 values obtained in this study (0.423 for Pb, 0.553 for Cd, and 0.584 for Hg) were lower than those reported by Bhagat et al. (2021). This difference is expected because the studies differ substantially in terms of environmental media (river water versus sediments and soil), dataset characteristics, predictor variables, sample size, and modelling framework. Furthermore, the present study relied on five water quality parameters (pH, BOD, COD, TDS, and TSS) as predictors. It generated future inputs using simulated values based on the Compound Annual Growth Rate (CAGR), whereas previous studies used different datasets and preprocessing techniques. Similarly, Sharma et al. (2024) demonstrated that Random Forest effectively modelled non-linear relationships in water quality prediction by achieving the highest classification accuracy among the evaluated algorithms. Although Sharma et al. addressed a classification problem rather than a regression task, the findings further support the suitability of ensemble learning methods for environmental applications. Overall, the comparison indicates that the relative superiority of XGBoost observed in this study is consistent with the broader literature. At the same time, differences in absolute performance can reasonably be attributed to variations in study design, data availability, environmental conditions, and prediction objectives.

The predictive performance of the Random Forest and XGBoost models is affected by several limitations in the available dataset, which also contribute to the observed prediction errors. First, the historical dataset is relatively small, consisting of only 324 data points collected between 2015 and 2023, thereby limiting the model's ability to capture complex temporal patterns and rare events that affect river water quality. Second, missing data in some monitoring periods required data preprocessing, which may have introduced uncertainty in the predictor variables. Third, these models were developed using only five water quality parameters (pH, BOD, COD, TDS, and TSS) as predictor variables. However, heavy metal concentrations are also influenced by other environmental and anthropogenic factors, such as rainfall intensity, river discharge, industrial waste, land-use changes, and seasonal variations, which were not included due to limited data availability. Fourth, the study focused on evaluating and comparing the predictive performance of Random Forest and XGBoost models for forecasting Pb, Cd, and Hg concentrations. Therefore, feature importance and model explainability analyses were outside the scope

of this research. Consequently, the relative contributions of the individual predictor variables (pH, BOD, COD, TDS, and TSS) to the model predictions were not assessed.

Fifth, the forecasting results are presented as point estimates without explicit quantification of prediction uncertainty. Therefore, the predicted heavy metal concentrations should be interpreted as expected values rather than precise future outcomes. Sixth, feature importance and model explainability were not evaluated in the present study. Consequently, the relative contribution of pH, BOD, COD, TDS, and TSS to the prediction of Pb, Cd, and Hg could not be quantitatively established. Also, hyperparameter optimisation was not performed for either Random Forest or XGBoost. The models were developed using predefined hyperparameters, with 100 estimators for both algorithms, while other parameters were set to their default values. In addition, formal statistical significance testing was not performed to determine whether the differences in predictive performance between Random Forest and XGBoost were statistically significant. The models were compared descriptively using MAE, MAPE, MSE, RMSE, and R^2. Therefore, the lower error values and higher R^2 obtained by XGBoost indicate better observed performance but do not necessarily represent statistically significant differences.

These factors collectively explain the residual prediction errors and indicate that, although the models provide useful predictive capability, their performance could be improved by incorporating additional explanatory variables, longer monitoring periods, and larger datasets in future studies.

DISCUSSION

The predicted behaviour of Pb, Cd, and Hg should be interpreted in light of the environmental and anthropogenic characteristics of the Ciliwung River watershed. As the river flows through highly urbanised areas of Jakarta and its surrounding regions, changes in land use and intensive human activities may influence the transport and accumulation of contaminants. Industrial and commercial activities may contribute heavy metals through wastewater discharge, surface runoff, and other anthropogenic inputs. At the same time, densely populated urban areas may increase pollutant loading through domestic wastewater, traffic-related emissions, drainage systems, and stormwater runoff. These sources can vary spatially and temporally along the river, potentially contributing to fluctuations in heavy metal concentrations.

Seasonal and hydrological conditions may also influence the observed and predicted concentrations. During periods of high rainfall, increased surface runoff can transport contaminants accumulated on urban surfaces and soils into the river. Increased river discharge may subsequently alter contaminant dilution and transport. Conversely, during periods of lower flow, reduced dilution may contribute to relatively higher concentrations of dissolved pollutants. The interaction between rainfall, runoff, river discharge, and sediment processes can therefore affect the temporal distribution of heavy metals. However, because rainfall, river discharge, and seasonal variables were not included as predictors in the present model, their effects cannot be quantitatively isolated from the model predictions.

The relatively stable predicted concentrations of Pb, Cd, and Hg during the 2024–2030 forecasting period may indicate that, under the simulated future conditions, the model does not anticipate substantial changes in overall water-quality conditions, as represented by the available historical data. However, this should not be interpreted as evidence that pollutant inputs will remain constant. Future changes in industrial activity, urban expansion, land use, rainfall patterns, wastewater management, and river hydrology could alter heavy metal concentrations in ways that the current model does not capture. Therefore, continued monitoring is necessary to determine whether future environmental changes deviate from the simulated conditions used in this study.

Land-use characteristics within the Ciliwung watershed may further influence pollutant inputs. Areas dominated by residential, commercial, industrial, and densely developed land uses can generate different pollutant loads and runoff characteristics. In addition, heavy metals can be associated with suspended solids and sediments, meaning that changes in TSS and river hydrodynamics may affect their transport and distribution. The relationships between the selected predictors (pH, BOD, COD, TDS, and TSS) and heavy metal concentrations may therefore reflect, in part, the combined influence of wastewater inputs, organic matter, suspended material, and other watershed processes. Nevertheless, the present study does not directly establish causal relationships between individual environmental factors and heavy metal concentrations.

The presence of Pb, Cd, and Hg in the Ciliwung River is primarily associated with anthropogenic activities occurring throughout the watershed. Lead contamination stems from industrial waste disposal, urban runoff, vehicle emissions, and improper solid waste disposal. Cadmium contamination primarily stems from the production of plastics, fertilisers, and industrial wastewater [20]. Mercury contamination is caused by industrial waste and medical waste. In addition, domestic wastewater and surface runoff during rainfall events can transport these contaminants into the river, contributing to the accumulation and redistribution of heavy metals. These pollution sources are consistent with the characteristics of the Ciliwung watershed, which flows through densely populated urban areas with diverse residential, commercial, and industrial activities. Consequently, the predicted heavy metal concentrations reflect the combined influence of these anthropogenic inputs and the complex environmental processes governing contaminant transport within the river system.

In the United States, water treatment research has shifted beyond technological innovation to include infrastructure modernisation, addressing ageing water networks that increase the risks of leaks, corrosion, and heavy metal contamination. Current efforts emphasise artificial intelligence, predictive management, and optimised treatment operations to improve efficiency and resilience through proactive rather than reactive water management [50].

Climate change adaptation is another major priority, with research focusing on wastewater reuse, stormwater management, and circular water economy approaches. Wastewater is increasingly regarded as a valuable resource, supported by standards, risk-management frameworks, and policy systems to ensure sustainable water supplies under climate uncertainty [51].

Water treatment research has also evolved into a socio-technical field that integrates environmental justice, equitable access to clean water, and data-driven governance using artificial intelligence, IoT, and predictive modelling. These developments position water treatment as a foundation for environmental sustainability, infrastructure resilience, public health, and long-term water security [52].

Overall, leading countries increasingly integrate technological innovation, resilient infrastructure, and sustainable water governance into comprehensive water security systems, making water management a key pillar of sustainable development.

Using the search terms "Ciliwung River Heavy Metal" on the Scopus website, 13 papers were found that reported research on this topic, which can be analysed in in table 3.

1. Research Trend (Documents by Year)

The publication trend remained low and sporadic from the mid-1990s to around 2015, indicating limited research attention. Since 2016, output has become more consistent, with a sharp increase after 2020 and a peak in the most recent year. This growth reflects increasing scientific interest, driven by environmental concerns, urban pollution, and stricter regulations. It also reflects advances in analytical technologies, marking the field's transition from an emerging topic to an established area of research.

2. Documents by Affiliation

The affiliation analysis shows that Indonesian universities and research institutions dominate publication output, with institutions such as Universitas Indonesia, Universitas Gadjah Mada, and IPB contributing most frequently. This result reflects strong domestic engagement with national environmental issues. It suggests that expertise is distributed across multiple institutions rather than concentrated in a single research centre, supporting collaborative and multidisciplinary research.

3. Documents by Country or Territory

The country distribution is dominated by Indonesia, with Australia contributing a much smaller share, likely through international collaboration. This pattern reflects the location-specific nature of the research, which is closely linked to Indonesian river systems and environmental challenges. The limited involvement of other countries highlights opportunities for broader international collaboration, particularly in comparative studies of river pollution and urban water sustainability.

4. Documents by Subject Area

The subject-area distribution demonstrates the field's interdisciplinary nature. Environmental Science is the dominant discipline, followed by Agricultural and Biological Sciences, reflecting research on water quality, pollution, and ecological impacts. Contributions from Chemistry and Biochemistry emphasise contaminant analysis, while Computer Science and Social Sciences highlight growing interest in

modelling, data analysis, policy, and governance. Overall, the field integrates natural sciences, analytical methods, and socio-environmental perspectives.

5. Documents by Type

The document type distribution is dominated by journal articles, accounting for over 90% of publications, while conference papers comprise only a small proportion. This result indicates that the field is well established, with research primarily disseminated through peer-reviewed journals, reflecting an emphasis on scientific rigour, reproducibility, and policy-relevant findings.

6. Documents by Funding Sponsor

Funding is primarily provided by Indonesian government agencies and universities, with international institutions playing a smaller role. This result reflects strong domestic commitment to environmental pollution and water resource management. At the same time, the limited international funding highlights opportunities to expand collaboration through multinational research grants and global environmental programmes.

Collectively, the figure depicts a rapidly growing, interdisciplinary, and policy-relevant research field with strong national engagement. The increase in publications, active institutional participation, and government support reflects rising awareness of environmental challenges and the importance of river management. At the same time, opportunities remain to strengthen international collaboration, comparative studies, and advanced analytical approaches to enhance the field's global impact.

Table 4
Scopus results for the Ciliwung River heavy metal search, for each Title, Author, and Method.

No.	Title	Author	Content	Method
1	Intestinal bacterial diversity of suckermouth catfish (Pterygoplichthys pardalis) in the Cd, Hg, and Pb-contaminated Ciliwung River, Indonesia.	Elfidasari, D., Rijal, M. S., Shalsabilla, S. E., Fadila, D. S. R., Cici, A., Pikoli, M. R., & Sugoro, I. (2023). [53]	Intestinal bacteria enable P. pardalis survival in heavily metal-polluted Ciliwung River; NGS reveals a diverse core microbiota negatively correlated with contaminants, highlighting fish adaptability and the potential for bioremediation of contaminated river sediments.	The study used field sampling, heavy metal analysis of water and sediments, and NGS-based gut microbiome sequencing of *P. pardalis*, with microbial diversity statistically correlated to Cd, Hg, and Pb contamination.
2	Heavy metal levels in fish products in Indonesia: a survey	Yulyana, A., Hastuti, A. A. M. B., Rohman, A., Setiawan, B., & Khasanah, F. (2023). [54]	Indonesia's fisheries face heavy metal risks; arsenic, cadmium, mercury, and lead exceed limits in polluted waters, requiring stronger surveillance, stricter regulation, and fishing restrictions to protect consumer food safety nationwide.	The method is a systematic literature review that compiles published data on arsenic, cadmium, mercury, and lead concentrations in Indonesian fish, and compares them with national food safety regulations and official surveillance reports.
3	Assessment of lead heavy metal pollution in the Ciliwung River.	Juwana, I., Nurjayati, R., Maria, R., Santoso, H., & Marganingrum, D. (2023, November). [55]	Survey-based assessment measured lead in Ciliwung River water and sediments using in situ physicochemical parameters and AAS analysis, revealing anthropogenic pollution exceeding standards, downstream degradation, and implications for the sustainability of drinking-water treatment.	The study applied field surveys and analytical methods, collecting water and sediment samples, measuring physicochemical parameters in situ, and determining lead concentrations using Atomic Absorption Spectrophotometry to assess river pollution.
4	Multiplex PCR detection of mackerel-based food adulteration with pleco and	Widayanti, R., Nugroho, H. A., Megarani, D. V., Widiasih, D. A., &	Multiplex PCR and DNA barcoding detected pleco and chicken adulteration in mackerel products	The study applied DNA barcoding and species-specific multiplex PCR to analyse 21 processed

No.	Title	Author	Content	Method
	chicken in selected areas around the Ciliwung River, Indonesia	Pakpahan, S. (2023). [56]	near the Ciliwung River, revealing significant food fraud risks, altered nutrition, and potential heavy metal contamination threatening food safety.	mackerel products, amplifying diagnostic fragments and constructing phylogenetic trees to detect pleco and chicken adulteration.
5	Cadmium and lead uptake by the wild swamp eel in the populous river	Nurhasanah, N., Sulistyowati, L., Riani, E., & Cordova, M. R. (2023). [57]	The study assessed cadmium and lead accumulation in Ciliwung River sediments and eel organs using EPA Method 3051A digestion and intake estimation, revealing downstream pollution, bioaccumulation risks, and implications for food safety.	The research used field sampling of sediments and eels across river sections, EPA 3051a digestion for metal analysis, quality-controlled laboratory measurements, and dietary intake estimation to evaluate cadmium and lead bioaccumulation.
6	Identification of Mackerel (Scomberomorus commerson) and Suckermouth Catfish (Pterygoplichthys sp.) in Otak-otak	Suseno, D., & Razari, I. (2023). [58]	PCR and DNA sequencing authenticated mackerel and suckermouth catfish in otak-otak, while AAS measured heavy metals, revealing frequent catfish adulteration and lead contamination exceeding safety limits.	The method combined PCR with species-specific primers and Sanger sequencing to authenticate fish species in otak-otak, while Atomic Absorption Spectrophotometry quantified heavy metal contamination in adulterated samples.
7	Survivability of Pterygoplichthys pardalis from Ciliwung River, Jakarta, Indonesia, based on metal toxicity test and bioaccumulation of cadmium, mercury, and lead	Elfidasari, D., Sugoro, I., Nabilah, D., & Amalia, S. N. (2022). [59]	The study evaluated the survivability of Pterygoplichthys pardalis under Cd, Hg, and Pb exposure using acute toxicity tests, mortality assessments, and bioaccumulation analyses, demonstrating high tolerance in polluted freshwater environments.	The method involved laboratory-based acute toxicity experiments exposing P. pardalis to Cd, Hg, and Pb, measuring lethal thresholds, mortality rates, and metal bioaccumulation to assess survivability in polluted freshwater conditions.
8	Heavy Metal Uptake Test by Aquatic Plants Tissue	M. F. I. Nugraha, R. Yunita, A. Julzarika, T. A.	Aquarium experiments tested tissue-cultured	Using Neutron Activation Analysis (NAA) for accurate

No.	Title	Author	Content	Method
	Culture Products with Neutron Activation Analysis in Ciliwung River	Putra, and T. R. Mulyaningsih (2021).[60]	aquatic plants for phytoremediation in Ciliwung River water, using periodic water replacement and Neutron Activation Analysis to quantify heavy-metal uptake and plant survival.	measurement of heavy metals.
9	Heavy metal concentration in water, sediment, and pterygoplichthys pardalis in the Ciliwung River, Indonesia	Elfidasari, D., Ismi, L. N., & Sugoro, I. (2020). Heavy metal concentration in water, sediment, and pterygoplichthys pardalis in the Ciliwung River, Indonesia. AACL Bioflux, 13(3), 1764–1778. [61]	XRF analysis quantified heavy metals in Ciliwung River water, sediments, and Pterygoplichthys pardalis along upstream–downstream gradients, revealing regulatory exceedances and substantial sediment–fish bioaccumulation correlations.	The method involved systematic sampling of water, sediments, and P. pardalis along the Ciliwung River, followed by X-Ray Fluorescence analysis to quantify heavy metals and assess bioaccumulation relationships.
10	Sources and drivers of contamination along an urban tropical river (Ciliwung, Indonesia): Insights from microbial DNA, isotopes and water chemistry	Duvert, C., Priadi, C. R., Rose, A. M., Abdillah, A., Marthanty, D. R., Gibb, K. S., & Kaestli, M. (2019). [62]	Integrated isotopic, chemical, and microbial tracer analyses, along with longitudinal river sampling and Bayesian source tracking, were used to identify spatial–temporal contamination pathways, hotspots, and hydrological controls in the urban Ciliwung River system.	The method combined longitudinal river sampling with isotopic, chemical, and microbial tracer measurements, seasonal comparison, and Bayesian microbial source tracking to identify contamination pathways, sources, and hotspots in an urban river system.
11	The Correlation Between Heavy Metal and Nutrient Content in Plecostomus (Pterygoplichthys pardalis) from Ciliwung River in Jakarta	Elfidasari, D., Ismi, L. N., Shabira, A. P., & Sugoro, I. (2018). [63]	The study analysed heavy metals and nutritional composition of plecostomus flesh using XRF, Kjeldahl protein analysis, and fat extraction, revealing that metal contamination affects protein and fat levels across fish sizes.	The method involved sampling plecostomus from the Ciliwung River, measuring heavy metals by XRF, determining protein by Kjeldahl analysis, determining fat by extraction, and comparing nutritional content across fish size classes.

No.	Title	Author	Content	Method
12	The analysis of the concentration of heavy metals cadmium, mercury and lead in the flesh of suckermouth catfish (Pterygoplichthys pardalis) in Ciliwung River, Indonesia	Ernawati, Y. (2014). [64]	Heavy metal analysis of suckermouth catfish from the Ciliwung River showed cadmium and mercury below limits. Still, lead exceeded standards, indicating consumption risks for fish-based foods from Bogor, Depok, and Jakarta.	The method involved sampling Pterygoplichthys pardalis from Bogor, Depok, and Jakarta, followed by laboratory determination of cadmium, mercury, and lead concentrations in fish flesh using Atomic Absorption Spectrophotometry.
13	River water quality study in the vicinity of Jakarta	Palupi, K., Sumengen, S., Inswiasri, S., Agustina, L., Nunik, S. A., Sunarya, W., & Quraisyn, A. (1995). [65]	Field studies using the Water Quality Index and heavy metal analysis assessed the Ciliwung, Sunter, and Krukut Rivers across seasons, revealing poor water quality and pollution despite most metals meeting Indonesian standards.	The method involved seasonal field sampling of river water at upstream, intake, and estuary sites, calculating Water Quality Index scores and measuring heavy metal concentrations to assess urban river water quality.
14	A Review of Analysing Heavy Metal Pollution with Data Science Approaches. In International Conference on Trends in Sustainable Computing and Machine Intelligence (pp. 296–311). Cham: Springer Nature Switzerland	Gozali, L., & Angelica, C. E. (2025, September). [66]	Reviews relationships between electric vehicle adoption and carbon dioxide emissions, identifying research gaps in forecasting, correlations, geographic focus, data limitations, and comparing existing findings on EV environmental impacts.	Systematic literature review using ScienceDirect, filtering 3,877 articles by relevance, selecting eight studies, comparing forecasting models, variables, regions, and methodologies to identify gaps in EV–CO_2 emission research.
15	Greenhouse gas emissions forecasting analysis in Jakarta towards net-zero emissions by 2050.	Gozali, L., Liu, K. V., Adianto, A., Nasution, S. R., & Doaly, C. O. (2025). [67][68]	The study analyses historical greenhouse gas emissions, forecasts future trends, and evaluates policy scenarios to guide pathways toward achieving net-zero emissions by 2050 at the national and sectoral levels, using empirical data.	Research applies time-series forecasting models with scenario analysis, using emissions inventory data to project future trajectories, compare mitigation pathways, and assess progress toward net-zero targets under alternative assumptions.

Table 4 summarises 15 studies that collectively provide a comprehensive understanding of heavy metal pollution, ecological responses, food safety, and environmental management in the Ciliwung River. These studies can be grouped into four themes: (1) contamination assessment and environmental monitoring, (2) bioaccumulation and ecological adaptation, (3) food safety and public health, and (4) emerging data-driven environmental governance.

The first theme focuses on contamination assessment. Early studies established baseline water quality conditions, while later research employed advanced techniques such as Water Quality Index analysis, Atomic Absorption Spectrophotometry (AAS), isotope tracing, and microbial source tracking. Collectively, these studies demonstrate that the Ciliwung River experiences cumulative pollution from industrial, domestic, and urban activities, reflecting a shift from conventional monitoring toward integrated system-level environmental diagnostics.

The second theme examines bioaccumulation and ecological adaptation. Research on aquatic organisms, particularly *Pterygoplichthys pardalis*, shows that pollution-tolerant species accumulate heavy metals while developing physiological and microbial adaptations that enable them to survive in contaminated environments. These findings highlight the potential role of invasive species as bioindicators and candidates for bioremediation.

The third theme addresses food safety and public health. Studies consistently report heavy metal contamination in fish and seafood products (Yulyana et al., 2023; Widayanti et al., 2023; Suseno & Razari, 2023), with some exceeding food safety limits. These findings demonstrate that river pollution threatens not only aquatic ecosystems but also food security and human health. Related research on phytoremediation further reflects growing interest in sustainable, nature-based remediation strategies.

The final theme represents a recent shift toward predictive environmental governance (Gozali & Angelica, 2025; Gozali et al., 2025). Recent review and forecasting studies integrate data science, predictive modelling, and policy analysis, indicating that river pollution research is evolving from environmental monitoring to proactive environmental management that supports sustainable urban planning and climate adaptation.

Overall, the literature demonstrates a clear progression over the past three decades, from pollution identification to ecological impact assessment and, more recently, to predictive and interdisciplinary approaches integrating molecular biology, analytical chemistry, and data science. Together, these studies position the Ciliwung River as a model system for understanding the interactions among heavy metal contamination, ecosystem health, food safety, and sustainable watershed management.

The comparison matrix in Table 4 further shows that existing studies are predominantly descriptive and retrospective, focusing on measuring contamination levels, identifying pollution sources, assessing bioaccumulation, and evaluating ecological or human health risks using established analytical techniques such as AAS, XRF, NAA, PCR, NGS, and isotopic tracers. Although these methods provide

robust evidence of current environmental conditions, they offer limited capability to anticipate future pollution trends.

In contrast, this study advances the current state of the art by applying ensemble machine-learning methods, namely XGBoost and Random Forest, to forecast future heavy metal concentrations. Rather than merely assessing existing contamination, the proposed approach enables early warning, scenario analysis, and proactive environmental management. To the best of our knowledge, previous studies on the Ciliwung River have not applied or compared these advanced forecasting models for predicting heavy metal pollution, highlighting the novelty of the present research.

CONCLUSION

This study demonstrates that the XGBoost algorithm achieves better predictive performance than Random Forest in predicting the concentrations of lead (Pb), cadmium (Cd), and mercury (Hg) in the Ciliwung River. This result is evidenced by evaluations using MAE, MAPE, MSE, RMSE, and the coefficient of determination (R^2), where XGBoost consistently produces lower prediction errors and higher R^2 values. However, the R^2 values of 0.423 for Pb, 0.553 for Cd, and 0.584 for Hg indicate moderate predictive performance, suggesting that a substantial proportion of the variability in heavy metal concentrations remains unexplained. The XGBoost predictions for the 2024–2030 period estimate Pb concentrations of 14.285–19.483 µg/L, Cd concentrations of 7.413–9.998 µg/L, and Hg concentrations of 0.227–0.457 µg/L. Under the simulated future conditions, the projected Pb, Cd, and Hg concentrations remain below the applicable Class II water quality standards for 2024–2030. However, these projections should not be interpreted as definitive evidence that future river water quality will remain safe, as unforeseen environmental and anthropogenic changes may affect heavy metal concentrations. Nevertheless, these results should be interpreted with caution, as the model has moderate explanatory power and does not account for all environmental and anthropogenic factors that may influence heavy metal concentrations. Therefore, the XGBoost model should be considered a supporting tool for water quality monitoring and assessment, rather than a substitute for direct laboratory measurements.

FUTURE RESEARCH

Justification for further research. Given consistent evidence of elevated metal loads, their documented accumulation in sediments and biota, and the socio-environmental vulnerability of populations living along the Ciliwung, focused research is needed that (a) maps spatial and seasonal variability with paired water–sediment–biota sampling, (b) determines metal speciation and bioavailability (including methylmercury), (c) quantifies source contributions using diagnostic indices and land-use/industrial inventories, and (d) links measured concentrations to human-health and ecological risk metrics. Such an integrated approach will provide the scientific basis for prioritised mitigation (industrial discharge control, sanitation upgrades, sediment management) and for monitoring programs that protect both local communities and Jakarta's coastal environment.

Further research is needed on the things that need to be paid attention to, namely strategic decisions that must be taken to create strength and resilience in the scientific, science and technology fields and be able to create practical applications in the field of environmental management, in improving the lives of the people of Jakarta to be better and more prosperous.

Future research should incorporate additional environmental and anthropogenic variables to improve the forecasting performance and generalizability of machine learning models. In particular, rainfall, river discharge, seasonal and hydrological conditions, land-use changes, industrial discharge records, wastewater inputs, and remote sensing data could be integrated with the existing water quality parameters. These variables may provide additional information on pollutant transport, dilution, runoff, and anthropogenic pressures that pH, BOD, COD, TDS, and TSS alone do not capture. The integration of high-resolution spatial and temporal data could therefore improve the model's ability to capture complex and non-stationary patterns of heavy metal contamination in the Ciliwung River.

Future studies should also incorporate feature importance and explainability techniques, such as permutation importance and SHAP (SHapley Additive exPlanations), to identify the relative contribution of individual predictors and provide deeper insight into the environmental factors influencing Pb, Cd, and Hg concentrations. In addition, systematic hyperparameter optimisation and external validation using independent observations from different monitoring periods or locations should be conducted to improve model robustness and generalizability.

Further research could also develop an intelligent, real-time monitoring system by integrating IoT-based water-quality sensors with machine learning models such as XGBoost. Such a system could support continuous monitoring, early detection of anomalous heavy metal concentrations, and proactive environmental management. Long-term field observations should be used to continuously validate and update the models, particularly under changing environmental and urban conditions.

Furthermore, there must be a long-term validation system that uses observation data collected in the new main field, which is essential for evaluating and validating the model's stability and adaptation using robust machine learning.

PRACTICAL AND SOCIAL IMPLICATIONS

In terms of practical implications, this study suggests that the machine learning approach, specifically XGBoost, can provide an effective predictive model for monitoring river pollution from existing heavy metals using continuously and routinely available water quality parameters. This alternative model provides greater efficiency in terms of cost expenditure for related agencies compared to conducting continuous laboratory experiments, which generally involve several other elements, resulting in increased costs and requiring a long time and recording local conditions at specific observation points. Early predictions of heavy metal contamination, such as Pb, Cd, and Hg, are supported by more proactive environmental and pollution management, so that interested agencies can provide an overview of potential risk trends before exceeding the permissible threshold. The forecast results for 2025–2030

also provide future strategic policy guidance for future environmental management programs. This model can also optimise sampling frequency and prioritise high-risk locations for the design of preventive mitigation policy strategies in line with Government Regulation of the Republic of Indonesia no. 22 of 2021.

Furthermore, this research contributes significantly to strategic operational policies and environmental management, as well as to water resources in other large cities. This forecasting analysis can also help river basin governments, environmental change management agencies, and local and regional governments shift from reactive, compliance-based monitoring to proactive, predictive, and adaptive environmental management. Artificial intelligence-guided water quality evaluations encourage data-driven decision-making, enhance environmental monitoring capabilities, and increase the capacity of national and regional institutions to manage the pressures posed by polluted environments caused by industrialisation and urban expansion.

From the social implications, this research has a significant impact on improving public health and the sustainability of environmental welfare. Although the calculation shows that the influence of heavy metals remains within levels that do not exceed the thresholds of existing regulations, the future trend of heavy metal pollution, Pb and Cd, indicates a potentially dangerous buildup over a long period of time. Pb and Cd themselves pose a high risk in the future through the process of bioaccumulation in the environment or aquatic ecosystems and the food chain supply. Early identification of trends can help prevent public health interventions and reduce pollution and exposure to risk over a fairly long period. This condition also provides the community with knowledge about which behaviours and habits carry a high risk of environmental pollution. In addition, transparent environmental prediction knowledge can increase public trust and encourage community participation in efforts to conserve rivers and develop a collaborative environmental governance system among the central government, local governments, the industrial sector, and the community.

More comprehensively and integratively, this research, which uses artificial intelligence methods, bridges the gap between scientific research and the social benefits it delivers to the community. With knowledge of environmental management and the ability to understand complex environmental data, actionable information can be derived. This research aims to achieve the Sustainable Development Goals, encouraging entrepreneurs and industry players to be more responsible and to participate in managing the environment and urban ecosystems in a well-managed, resilient way. This condition is realised by integrating machine learning forecasting methods into the strategic policy framework for good environmental ecosystems. The new and adaptive methods developed are expected to represent technological progress in advancing the welfare of the global community, providing clean water of sufficient quality and adopting a socially transformative approach to maintaining the availability of water resources and improving the quality of the environment for future generations.

Declarations

Declaration of competing interest

s

Funding

"The author(s) received no specific financial support for the research, authorship, and/or publication of this article."

Author Contribution

Term DefinitionConceptualization(Lina Gozali) Ideas; formulation or evolution of overarching research goals and aimsMethodology(Adianto) Development or design of methodology; creation of modelsValidation(Togar) Verification, whether as a part of the activity or separate, of the overall replication/ reproducibility of results/experiments and other research outputsFormal analysis(Jose Reyes) Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study dataInvestigation(Christyfianie Elva Angelica)Conducting a research and investigation process, specifically performing the experiments, or data/evidence collectionResources(Wan Hee Cheng)Provision of study materials, reagents, materials, patients, laboratory samples, instrumentation, computing resources, or other analysis toolsWriting - Original Draft (Christyfianie Elva Angelica) Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)Writing - Review & Editing (Lina Gozali) Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre-or postpublication stagesSupervision(Vikas Kumar) Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team

Data Availability

All data supporting the findings of this study are available within the paper and its Supplementary Information.

Table 3

Tables 3 is available in the Supplementary Files section.

References

1. M. Mokarram, A. Saber, and V. Sheykhi, "Effects of heavy metal contamination on river water quality due to release of industrial effluents," J Clean Prod, vol. 277, p. 123380, Dec. 2020, doi:

10.1016/J.JCLEPRO.2020.123380.

2. B. Sarker, K. N. Keya, F. I. Mahir, K. M. Nahiun, S. Shahida, and R. A. Khan, "Surface and Ground Water Pollution: Causes and Effects of Urbanisation and Industrialisation in South Asia," Guigoz Sci Rev, no. 73, pp. 32–41, Jul. 2021, doi: 10.32861/sr.73.32.41.
3. World Health Organization, Safe management of wastes from health-care activities: a summary. Switzerland: World Health Organization, 2017.
4. Y. He, W. Wang, Y. Chen, J. Hua, C. Deng, and H. Li, "Source-sink response analysis of heavy metals and soil pollution assessment in non-ferrous metal industrial agglomeration areas based on decision unit," Science of the Total Environment, vol. 906, Jan. 2024, doi: 10.1016/j.scitotenv.2023.167437.
5. A. Daripa, L. C. Malav, D. K. Yadav, and S. Chattaraj, "Metal contamination in water resources due to various anthropogenic activities," Metals in Water: Global Sources, Significance, and Treatment, pp. 111–127, Jan. 2023, doi: 10.1016/B978-0-323-95919-3.00022-7.
6. X. Peng et al., "Source apportionment and heavy metal health risk (HMHR) quantification from sources in a southern city in China, using an ME2-HMHR model," Environmental Pollution, vol. 221, pp. 335–342, Feb. 2017, doi: 10.1016/J.ENVPOL.2016.11.083.
7. Y. Sun et al., "Spatial distribution prediction of soil heavy metals based on sparse sampling and multi-source environmental data," J Hazard Mater, vol. 465, Mar. 2024, doi: 10.1016/j.jhazmat.2023.133114.
8. H. Chen, Y. Teng, S. Lu, Y. Wang, and J. Wang, "Contamination features and health risk of soil heavy metals in China," Science of The Total Environment, vol. 512–513, pp. 143–153, Apr. 2015, doi: 10.1016/J.SCITOTENV.2015.01.025.
9. Z. Li, Z. Ma, T. J. van der Kuijp, Z. Yuan, and L. Huang, "A review of soil heavy metal pollution from mines in China: Pollution and health risk assessment," Science of The Total Environment, vol. 468–469, pp. 843–853, Jan. 2014, doi: 10.1016/J.SCITOTENV.2013.08.090.
10. W. Jadaa and H. Mohammed, "Heavy Metals – Definition, Natural and Anthropogenic Sources of Releasing into Ecosystems, Toxicity, and Removal Methods – An Overview Study," Journal of Ecological Engineering, vol. 24, no. 6, pp. 249–271, 2023, doi: 10.12911/22998993/162955.
11. C. Yao et al., "Heavy metal pollution in agricultural soils from surrounding industries with low emissions: Assessing contamination levels and sources," Science of the Total Environment, vol. 917, Mar. 2024, doi: 10.1016/j.scitotenv.2024.170610.
12. P. B. Angon et al., "Sources, effects and present perspectives of heavy metals contamination: Soil, plants and human food chain," Heliyon, vol. 10, no. 7, p. e28357, Apr. 2024, doi: 10.1016/J.HELIYON.2024.E28357.
13. S. Mitra et al., "Impact of heavy metals on the environment and human health: Novel therapeutic insights to counter the toxicity," J King Saud Univ Sci, vol. 34, no. 3, p. 101865, Apr. 2022, doi: 10.1016/J.JKSUS.2022.101865.

14. D. Kapoor and M. P. Singh, "Heavy metal contamination in water and its possible sources," Heavy Metals in the Environment: Impact, Assessment, and Remediation, pp. 179–189, Jan. 2021, doi: 10.1016/B978-0-12-821656-9.00010-9.
15. J. P. Vareda, A. J. M. Valente, and L. Durães, "Assessment of heavy metal pollution from anthropogenic activities and remediation strategies: A review," J Environ Manage, vol. 246, pp. 101–118, Sep. 2019, doi: 10.1016/J.JENVMAN.2019.05.126.
16. M. Adnan, B. Xiao, P. Xiao, P. Zhao, R. Li, and S. Bibi, "Research Progress on Heavy Metals Pollution in the Soil of Smelting Sites in China," Toxics, vol. 10, no. 5, pp. 1–30, 2022, doi: 10.3390/toxics10050231.
17. M. Aprilia, H. Effendi, and S. Hariyadi, "Water quality status based on Pollution Index and Water Quality Index of Ciliwung River, DKI Jakarta Province," IOP Conf Ser Earth Environ Sci, vol. 1109, no. 1, 2022, doi: 10.1088/1755-1315/1109/1/012051.
18. P. Yustitia, N. Islami, and N. Arina, "Study the Meaning Of Ciliwung River for Communities: Case Study Communities at Srengseng Sawah, South Jakarta)," Proceeding The 2nd ICHELSS, pp. 841–850, 2022.
19. P. Kumar et al., "Current assessment and future outlook for water resources considering climate change and a population burst: A case study of Ciliwung River, Jakarta City, Indonesia," Water (Switzerland), vol. 9, no. 6, 2017, doi: 10.3390/w9060410.
20. Dinas Lingkungan Hidup DKI Jakarta, “LAPORAN PEMANTAUAN KUALITAS AIR SUNGAI 2024,” Jakarta, 2024.
21. Pemerintah Republik Indonesia, Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup. Indonesia, 2021.
22. S. K. Bhagat, T. M. Tung, and Z. M. Yaseen, "Heavy metal contamination prediction using ensemble model: Case study of Bay sedimentation, Australia," J Hazard Mater, vol. 403, Feb. 2021, doi: 10.1016/j.jhazmat.2020.123492.
23. A. Sharma, R. Sharma, R. Rana, and A. Kalia, "Water quality prediction using Machine Learning Models," in E3S Web of Conferences, EDP Sciences, Nov. 2024. doi: 10.1051/e3sconf/202459601025.
24. R. Proshad et al., "Machine learning models with innovative outlier detection techniques for predicting heavy metal contamination in soils," J Hazard Mater, vol. 481, Jan. 2025, doi: 10.1016/j.jhazmat.2024.136536.
25. Liu, H., Chen, Z., Wang, Z., Xu, M., Wang, Y., Geng, J., & Yin, F. (2024). Water quality system informatics: An emerging inter-discipline of environmental engineering. Engineering, 43, 115–124.
26. Das, S. (2023). Introduction to water quality science. In An Introduction to Water Quality Science: Significance and Measurement Protocols (pp. 1–17). Cham: Springer International Publishing.
27. Qin, J., Ma, Z., & Li, Y. (2025). Review and outlook for research and development in aquatic life and ecosystems. Aquatic Life and Ecosystems, 1(1), 2.

28. Mutillod, C., Buisson, É., Mahy, G., Jaunatre, R., Bullock, J. M., Tatin, L., & Dutoit, T. (2024). Ecological restoration and rewilding: two approaches with complementary goals?. Biological Reviews, 99(3), 820–836.
29. M. Sharma and P. Kumar, "Environmental sources of heavy metals and their impacts on human health," Heavy Metal Toxicity and Neurodegeneration, pp. 317–326, Jan. 2025, doi: 10.1016/B978-0-443-36575-1.00030-3.
30. A. Hazrat, K. Ezzat, and Ilahi Ikram, "Environmental Chemistry and Ecotoxicology of Hazardous Heavy Metals: Environmental Persistence, Toxicity, and Bioaccumulation," J Chem, vol. 2019, no. Cd, pp. 1–14, 2019, [Online]. Available: https://doi.org/10.1155/2019/6730305
31. J. Briffa, E. Sinagra, and R. Blundell, "Heavy metal pollution in the environment and their toxicological effects on humans," Heliyon, vol. 6, no. 9, p. e04691, Sep. 2020, doi: 10.1016/J.HELIYON.2020.E04691.
32. K. H. Hama Aziz, F. S. Mustafa, K. M. Omer, S. Hama, R. F. Hamarawf, and K. O. Rahman, "Heavy metal pollution in the aquatic environment: efficient and low-cost removal approaches to eliminate their toxicity: a review," RSC Adv, vol. 13, no. 26, pp. 17595–17610, 2023, doi: 10.1039/d3ra00723e.
33. O. Oziegbe, E. J. Oziegbe, and O. Ojo-Omoniyi, "Bioremediation of heavy metals in aquatic environment: A review," Cleaner Chemical Engineering, vol. 11, p. 100193, Dec. 2025, doi: 10.1016/J.CLCE.2025.100193.
34. S. Ray and R. Vashishth, "From water to plate: Reviewing the bioaccumulation of heavy metals in fish and unraveling human health risks in the food chain," Emerg Contam, vol. 10, no. 4, p. 100358, Dec. 2024, doi: 10.1016/J.EMCON.2024.100358.
35. R. A. Bernhoft, "Mercury toxicity and treatment: A review of the literature," 2012, Hindawi Publishing Corporation. doi: 10.1155/2012/460508.
36. M.-T. Tsai, S.-Y. Huang, and S.-Y. Cheng, "Lead Poisoning Can Be Easily Misdiagnosed as Acute Porphyria and Nonspecific Abdominal Pain," Case Rep Emerg Med, vol. 2017, pp. 1–4, 2017, doi: 10.1155/2017/9050713.
37. M. Balali-Mood, K. Naseri, Z. Tahergorabi, M. R. Khazdair, and M. Sadeghi, "Toxic Mechanisms of Five Heavy Metals: Mercury, Lead, Chromium, Cadmium, and Arsenic," Apr. 13, 2021, Frontiers Media S.A. doi: 10.3389/fphar.2021.643972.
38. Jay. Heizer, Barry. Render, and Chuck. Munson, Operations management: sustainability and supply chain management. Pearson, 2016.
39. L. Seymour, P. J. Brockwell, and R. A. Davis, Introduction to Time Series and Forecasting., vol. 92, no. 440. 1997. doi: 10.2307/2965440.
40. A. Zemkoho, "A Basic Time Series Forecasting Course with Python," Operations Research Forum, vol. 4, no. 1, Mar. 2023, doi: 10.1007/s43069-022-00179-z.
41. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. New Jersey: John Wiley & Sons, 2015.

42. Gozali, L., Zulfan, P. M., Sukania, I. W., Ismail, S., Cheng, W. H., Masrom, M., & Robin, C. (2026). Implementation of Business Intelligence for Sales Prediction and Inventory Optimisation and Maintenance in SME Carton Box Company with Open-Source Application. *Management Systems in Production Engineering*, *34*(1).
43. G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning with Applications in R Second Edition, 2nd ed. Springer, 2021.
44. A. C. Müller and S. Guido, Introduction to Machine Learning with Python. Sebastopol: O'Reilly Media, 2016.
45. Giuseppe. Bonaccorso, Machine Learning Algorithms. Birmingham: Packt Publishing, 2017.
46. Gozali, L., Watun, A. M. L. O. U., Widodo, L., Ismail, S., Chofreh, A. G., Herwindiati, D. E., & Cheng, W. H. (2024). Designing orthopaedic seat cushion made of rattan using quality function deployment approach. Journal of Clinical Orthopaedics and Trauma, 58, 102771.
47. Kiparsky, M., Sedlak, D. L., Thompson Jr, B. H., & Truffer, B. (2013). The innovation deficit in urban water: the need for an integrated perspective on institutions, organisations, and technology. *Environmental engineering science*, *30*(8), 395–408.
48. Sivakumar, B. (2011). Global climate change and its impacts on water resources planning and management: assessment and challenges. *Stochastic Environmental Research and Risk Assessment*, *25*(4), 583–600.
49. Mandal, S., Yadav, A., Panwar, R., Kumar, S. S., Karthick, A., Priya, A., … Ganesh,S. S. (2025). Smart water management for SDG 6: a review of AI and iot-enabled solutions.*Water Conservation Science and Engineering*, *10*(2), 83.
50. Elfidasari, D., Rijal, M. S., Shalsabilla, S. E., Fadila, D. S. R., Cici, A., Pikoli,M. R., … Sugoro, I. (2023). Intestinal bacteria diversity of suckermouth catfish (Pterygoplichthys pardalis) in the Cd, Hg, and Pb contaminated Ciliwung River, Indonesia. Heliyon, 9(4).
51. Yulyana, A., Hastuti, A. A. M. B., Rohman, A., Setiawan, B., & Khasanah, F. (2023). Heavy metal levels in fish products in Indonesia: a survey. Food Research, 7(2), 74–84.
52. Juwana, I., Nurjayati, R., Maria, R., Santoso, H., & Marganingrum, D. (2023, November). Assessment of lead heavy metal pollution in Ciliwung River. In IOP Conference Series: Earth and Environmental Science (Vol. 1263, No. 1, p. 012033). IOP Publishing.
53. Widayanti, R., Nugroho, H. A., Megarani, D. V., Widiasih, D. A., & Pakpahan, S. (2023). Multiplex PCR detection of mackerel-based food adulteration with pleco and chicken in selected areas around Ciliwung River, Indonesia. Biodiversitas Journal of Biological Diversity, 24(3).
54. Nurhasanah, N., Sulistyowati, L., Riani, E., & Cordova, M. R. (2023). Cadmium and lead uptake by wild swamp eel in the populous river. Glob. J. Environ. Sci. Manag, 9, 497–514.
55. Suseno, D., & Razari, I. (2023). Identifikasi kandungan ikan tenggiri (Scomberomorus commerson) dan ikan sapu-sapu (Pterygoplichthys sp) pada otak-otak. Jurnal Pengolahan Hasil Perikanan Indonesia, 26(2), 191–205.

56. Elfidasari, D., Sugoro, I., Nabilah, D., & Amalia, S. N. (2022). Survivability of Pterygoplichthys pardalis from Ciliwung River, Jakarta, Indonesia based on metal toxicity test and bioaccumulation of cadmium, mercury, and lead. Aquaculture, Aquarium, Conservation & Legislation (AACL) Bioflux, 15(6), 3292–3302.
57. M. F. I. Nugraha, R. Yunita, A. Julzarika, T. A. Putra, and T. R. Mulyaningsih, "Heavy Metal Uptake Test by Aquatic Plants Tissue Culture Products with Neutron Activation Analysis in Ciliwung River", Int. J. Adv. Sci. Eng. Inf. Technol., vol. 11, no. 6, pp. 2323–2327, Dec. 2021.
58. Elfidasari, D., Ismi, L. N., & Sugoro, I. (2020). Heavy metal concentration in water, sediment, and pterygoplichthys pardalis in the Ciliwung River, Indonesia. AACL Bioflux, 13(3), 1764–1778.
59. Duvert, C., Priadi, C. R., Rose, A. M., Abdillah, A., Marthanty, D. R., Gibb, K. S., & Kaestli, M. (2019). Sources and drivers of contamination along an urban tropical river (Ciliwung, Indonesia): Insights from microbial DNA, isotopes and water chemistry. Science of the Total Environment, 682, 382–393.
60. Elfidasari, D., Ismi, L. N., Shabira, A. P., & Sugoro, I. (2018). The correlation between heavy metal and nutrient content in Plecostomus (Pterygoplichthys pardalis) from Ciliwung River in Jakarta. Biosaintifika: Journal of Biology & Biology Education, 10(3), 597–604.
61. Ernawati, Y. (2014). The analysis of the concentration of heavy metals cadmium, mercury and lead in the flesh of suckermouth catfish (Pterygoplichthys pardalis) in Ciliwung River, Indonesia. Aquaculture, Aquarium, Conservation & Legislation, 7(1), 33–42.
62. Palupi, K., Sumengen, S., Inswiasri, S., Agustina, L., Nunik, S. A., Sunarya, W., & Quraisyn, A. (1995). River water quality study in the vicinity of Jakarta. Water Science and Technology, 31(9), 17–25.
63. Gozali, L., & Angelica, C. E. (2025, September). A Review of Analysing Heavy Metal Pollution with Data Science Approaches. In *International Conference on Trends in Sustainable Computing and Machine Intelligence* (pp. 296–311). Cham: Springer Nature Switzerland
64. Liu, K. V., & Gozali, L. (2025, September). Achieving Net Zero: A Review of Greenhouse Gas Emission Forecasting Models and Approaches. In *International Conference on Trends in Sustainable Computing and Machine Intelligence* (pp. 498–508). Cham: Springer Nature Switzerland.
65. Liu, K. V., & Gozali, L. (2025, September). Achieving Net Zero: A Review of Greenhouse Gas Emission Forecasting Models and Approaches. In *International Conference on Trends in Sustainable Computing and Machine Intelligence* (pp. 498–508). Cham: Springer Nature Switzerland.

Figures

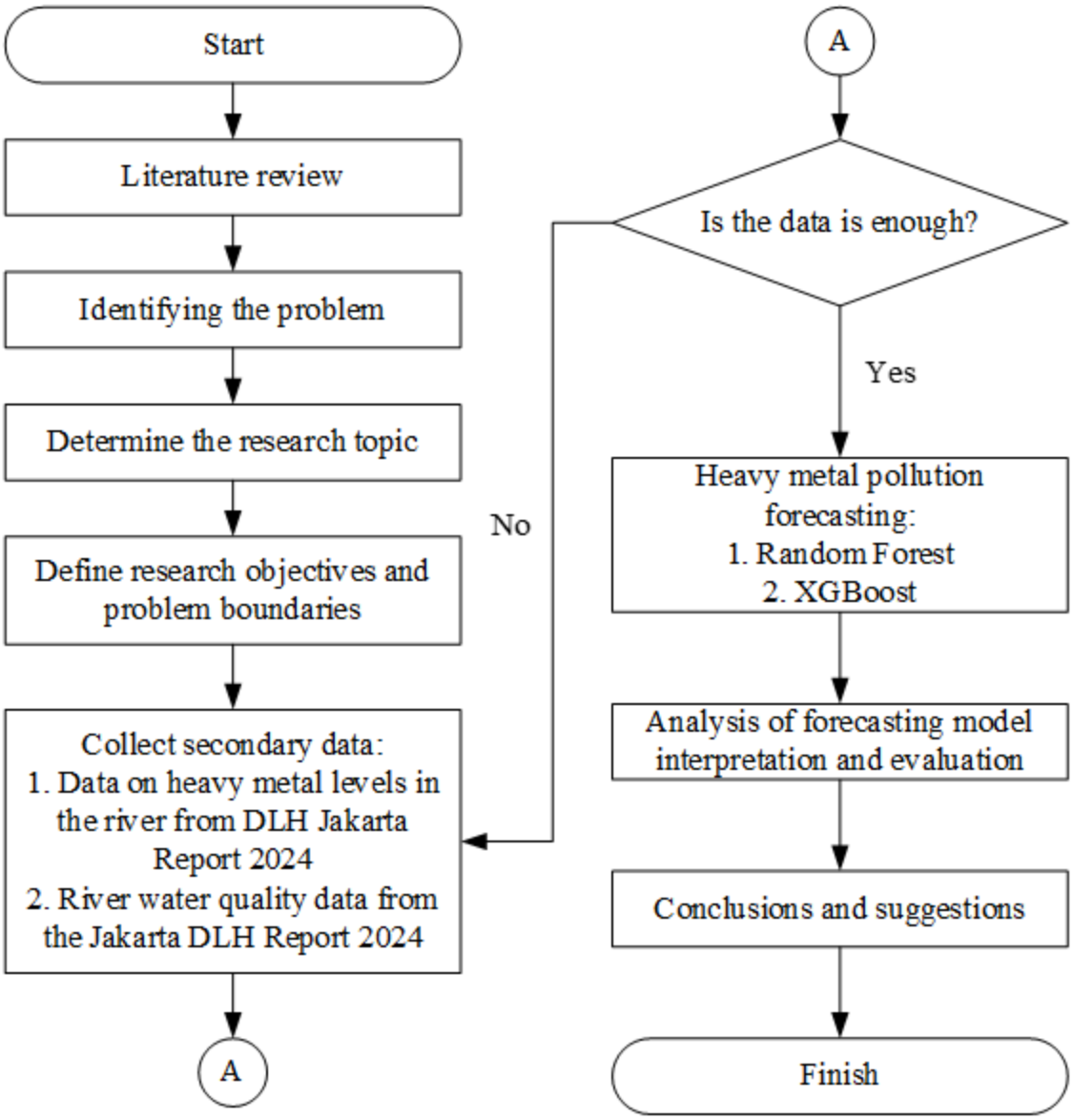

Figure 1

Research Methodology Flowchart

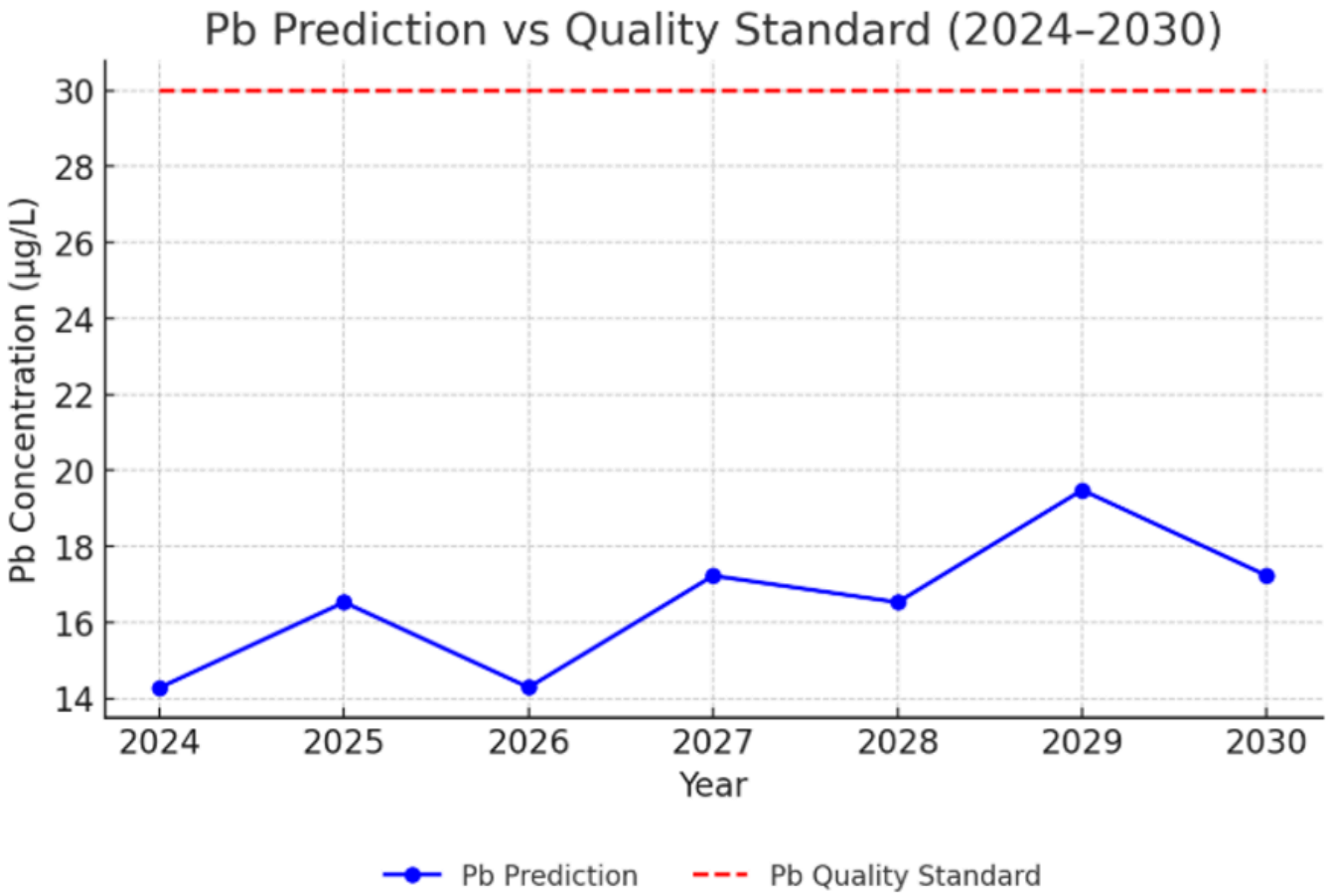

Figure 2

Pb Prediction for 2024-2030

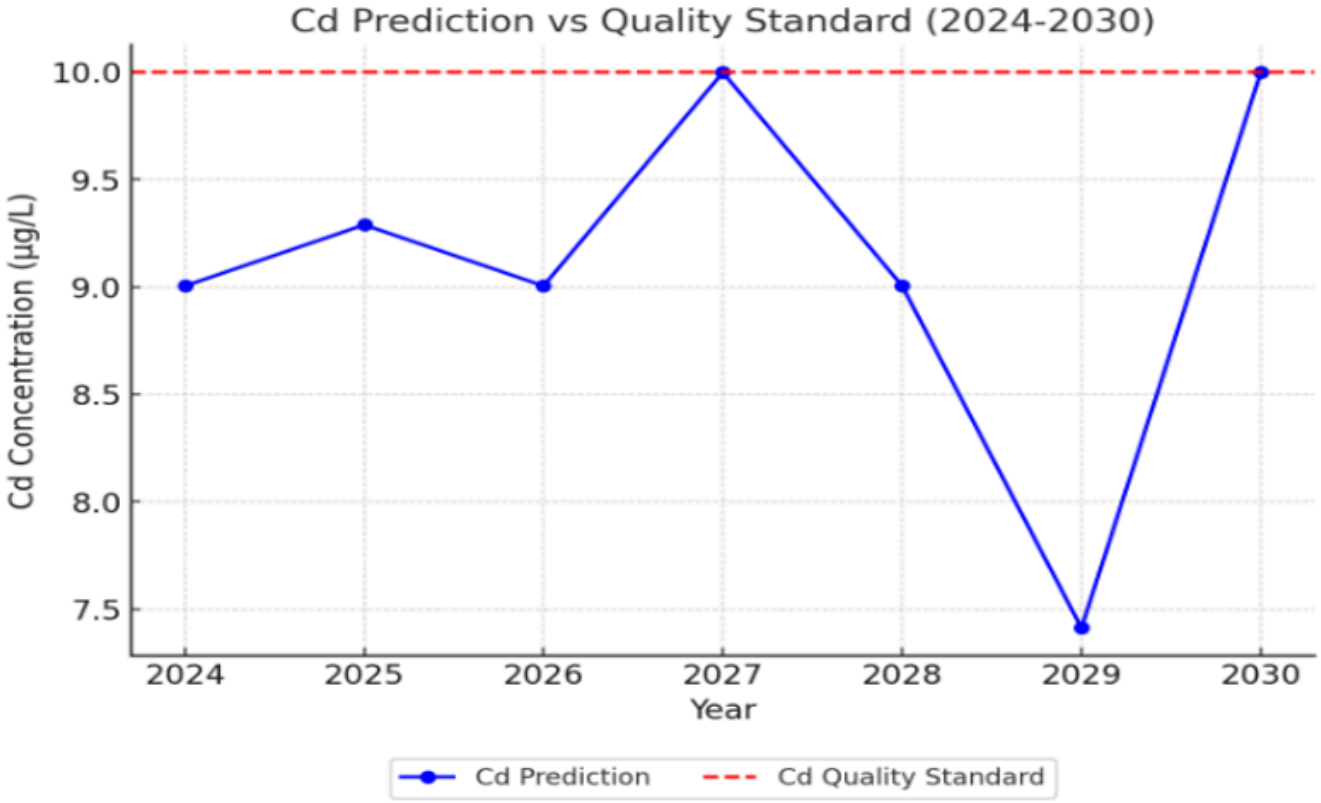

Figure 3

Cd Prediction for 2024-2030

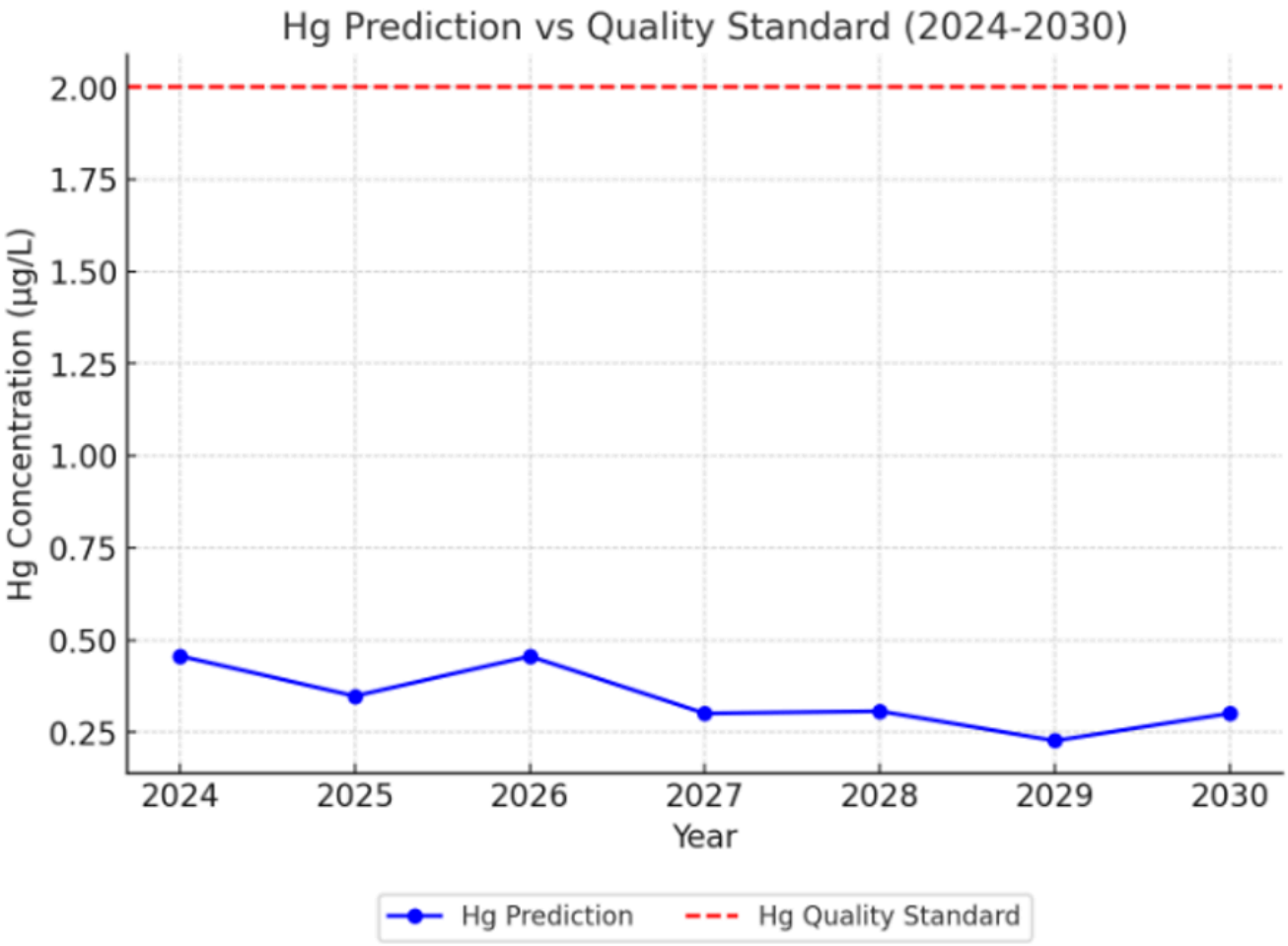

Figure 4

Hg Prediction for 2024-2030

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table3.docx